

DOCUMENT RESUME

ED 429 126

TM 029 669

AUTHOR Roberts, James S.; Donoghue, John R.; Laughlin, James E.
TITLE Estimating Parameters in the Generalized Graded Unfolding Model: Sensitivity to the Prior Distribution Assumption and the Number of Quadrature Points Used.
PUB DATE 1999-04-00
NOTE 34p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Attitude Measures; *Estimation (Mathematics); *Item Response Theory; Simulation
IDENTIFIERS Polytomous Items; *Unfolding Technique

ABSTRACT

The generalized graded unfolding model (J. Roberts, J. Donoghue, and J. Laughlin, 1998, 1999) is an item response theory model designed to unfold polytomous responses. The model is based on a proximity relation that postulates higher levels of expected agreement with a given statement to the extent that a respondent is located close to the statement on a unidimensional latent continuum. J. Roberts and others (1998) have examined the recovery of item and person parameters from the generalized graded unfolding model. Item and person parameters were estimated. This study used simulation methods to assess the sensitivity of model parameters to the prior distribution used in these estimation procedures. It also examined the effects of the number of quadrature points used in the numerical integration process and the calibration sample on the accuracy of the resulting parameter estimates. The results show that item parameter estimates derived from the marginal maximum likelihood procedure were fairly robust to discrepancies between the prior and true distributions of person parameters. Consequently, the person parameter estimates derived from the expected a posteriori procedure were generally robust as well, except for those individuals with the most extreme response patterns. The results also indicate that 20 quadrature points are adequate for the accurate recovery of model parameters. These findings will help establish the utility of the generalized graded unfolding model in applied measurement situations and will promote the use of item response theory in the attitude/preference measurement domain. (Contains 2 tables, 14 figures and 19 references.) (Author/SLD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

Estimating Parameters in the Generalized Graded Unfolding Model: Sensitivity to the Prior
Distribution Assumption and the Number of Quadrature Points Used

James S. Roberts
Medical University of South Carolina

John R. Donoghue
Educational Testing Service

James E. Laughlin
University of South Carolina

Paper Presented at the Annual Meeting of the National Council on Measurement in Education,
April 22, 1999, Montreal, PQ, Canada

Visit Our Web Site Devoted to Item Response Theory Models for Unfolding:
www.musc.edu/cdap/roberts

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
- ☒ Minor changes have been made to
improve reproduction quality.

- Points of view or opinions stated in this
document do not necessarily represent
official DERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY

James Roberts

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

TM029669

Abstract

The generalized graded unfolding model (Roberts, Donoghue and Laughlin, 1998, 1999) is an item response theory model designed to unfold polytomous responses. It is appropriate for data obtained in situations where subjects respond to a series of statements using either a binary or graded scale of agreement (e.g., situations where Thurstone or Likert attitude measurement procedures are employed). The model is based on a proximity relation which postulates higher levels of expected agreement with a given statement to the extent that a respondent is located close to the statement on a unidimensional latent continuum. Roberts et al. (1998) have examined the recovery of item and person parameters from the generalized graded unfolding model. Item parameters were estimated with a marginal maximum likelihood technique in which person parameters were numerically integrated out of the likelihood function based on a prior distribution. Person parameters were estimated using an expected a posteriori technique which also required numerical integration over a prior distribution of person parameters. Roberts et al. (1998) examined recovery of model parameters when the prior distribution of person parameters perfectly matched the true distribution. In contrast, this study used simulation methods to assess the sensitivity of model parameter estimates to the prior distribution used in these estimation procedures. It also examined the effects of the number of quadrature points used in the numerical integration process and the calibration sample size on the accuracy of the resulting parameter estimates. The results showed that item parameter estimates derived from the marginal maximum likelihood procedure were fairly robust to discrepancies between the prior and true distributions of person parameters. Consequently, the person parameter estimates derived from the EAP procedure were generally robust as well, except for those individuals with the most extreme response patterns. The results also indicated that 20 quadrature points were adequate for the accurate recovery of model parameters. These findings will help establish the utility of the generalized graded unfolding model in applied measurement situations and will promote the use of item response theory in the attitude/preference measurement domain.

Educational researchers typically use self-report questionnaires to assess attitudes toward or preferences for a variety of stimuli (e.g., attitude toward mathematics, preference for alternative types of instruction, etc.) Such questionnaires often contain a graded disagree-agree response format to gauge the level of individual agreement to a series of statements that range in content from negative, to neutral, to positive opinions. Several researchers (Andrich, 1996; Roberts, 1995; Roberts & Laughlin, 1996a, 1996b; Roberts, Laughlin & Wedell, 1999; Roberts, Wedell & Laughlin, 1998; van Schuur & Kiers, 1994) have argued that graded disagree-agree responses are generally more consistent with an unfolding model of the response process rather than the more popular cumulative model. Unfolding models are proximity models which imply that higher item scores, indicative of stronger levels of agreement, are more probable as the distance between an individual and an item on the underlying latent continuum decreases (Coombs, 1964). In this case, the underlying continuum will be characterized as a unidimensional, affective, bipolar continuum ranging from a very negative to a very positive orientation.

Roberts and colleagues (Roberts, 1995; Roberts, Donoghue & Laughlin, 1998, 1999; Roberts & Laughlin, 1996a, 1996b) have developed a family of item response theory models that implement an unfolding response mechanism. The most general of these models is called the Generalized Graded Unfolding Model (GGUM). At a conceptual level, the GGUM suggests that an individual will endorse an item to the extent that the sentiment conveyed by the item matches the individual's own opinion well. Psychometrically, the individual is expected to endorse the item to the extent that the individual is located close to the item on the latent continuum. The probability of obtaining an observed response Z_i under the GGUM is defined as:

$$Pr[Z_i = z | \theta_j] = \frac{\exp \left(\alpha_i [z (\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}] \right) + \exp \left(\alpha_i [(M-z)(\theta_j - \delta_i) - \sum_{k=0}^z \tau_{ik}] \right)}{\sum_{w=0}^C \left[\exp \left(\alpha_i [w (\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik}] \right) + \exp \left(\alpha_i [(M-w)(\theta_j - \delta_i) - \sum_{k=0}^w \tau_{ik}] \right) \right]} \quad (1)$$

for $z = 0, 1, \dots, C$; where θ_j is the location of the j th individual on the latent continuum, δ_i is the location of the i th item on the latent continuum, α_i is the discrimination parameter for the i th item, τ_{ik} is the k th subjective response category threshold for the i th item, C is the number of observable response categories minus 1, and M is equal to $2 \cdot C + 1$. Note that θ_j is an index of the j th individual's attitude and δ_i is an indicator of the i th item's affective content.

The GGUM yields single-peaked, bell-shaped response functions that imply higher levels of agreement to the extent that the individual and the item are close to each other on the latent continuum (i.e., to the extent that $|\theta_j - \delta_i|$ approaches zero). The GGUM is more general than other unfolding item response theory models in that it allows items to vary in their discrimination capabilities (via α_i), and it allows subjects to utilize the response scale differently for each item (via τ_{ik}).

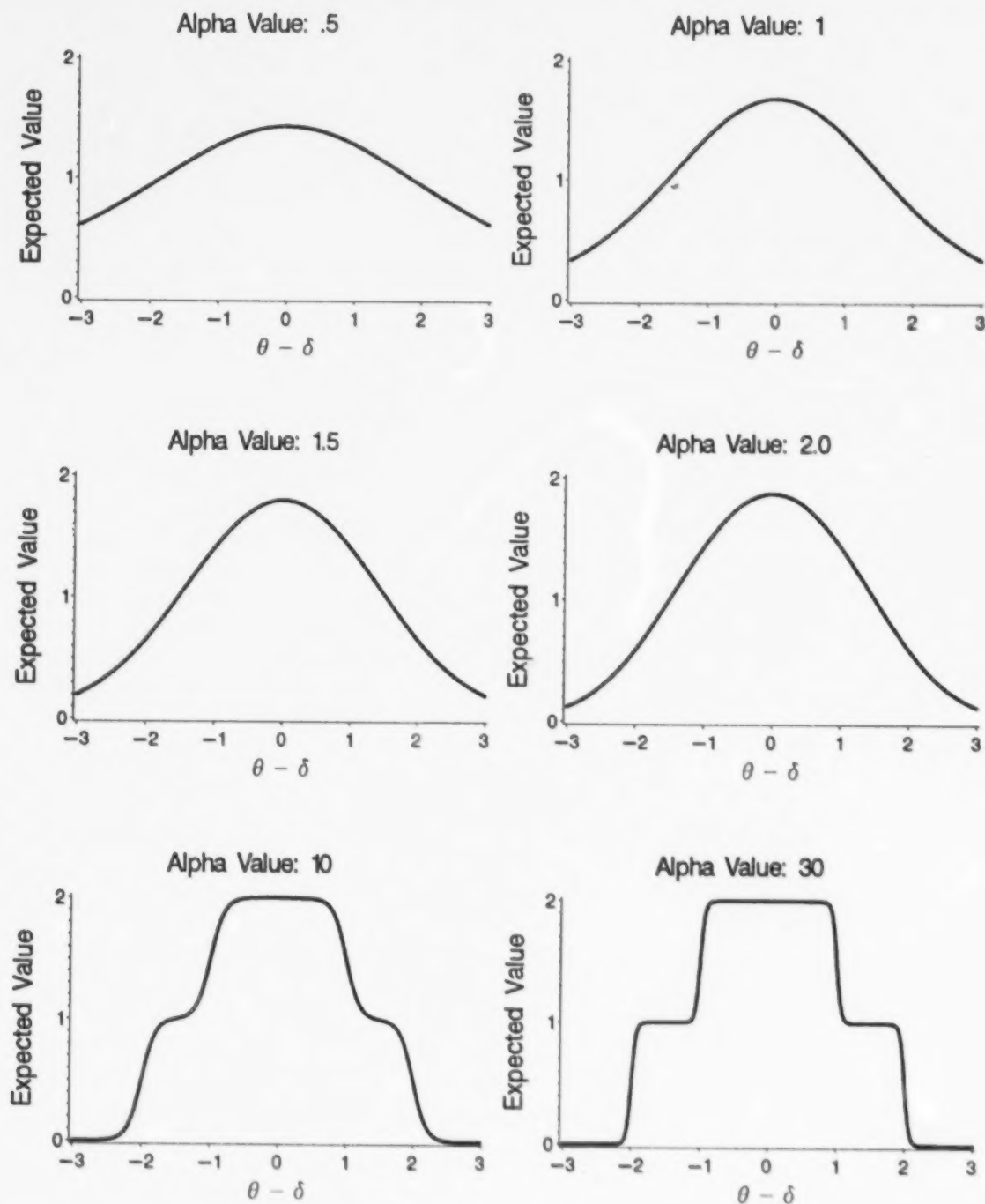
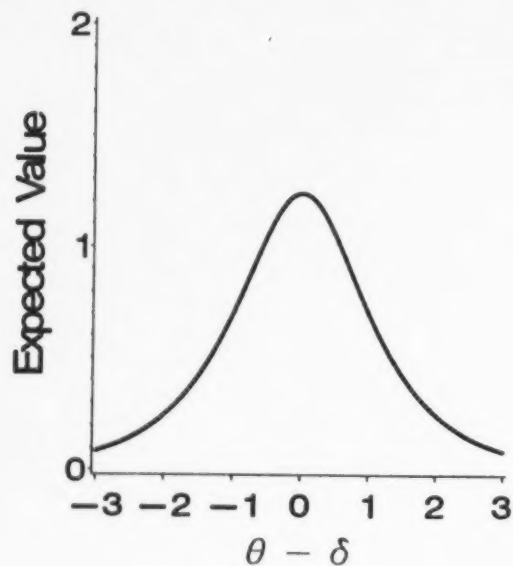
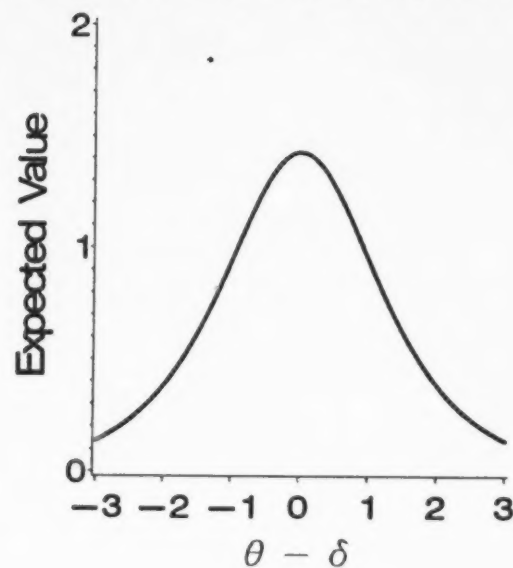


Figure 1. Response functions for a hypothetical 3-category item under the GGUM. Response functions are given for alternative values of the α parameter while the values for τ_a are held constant.

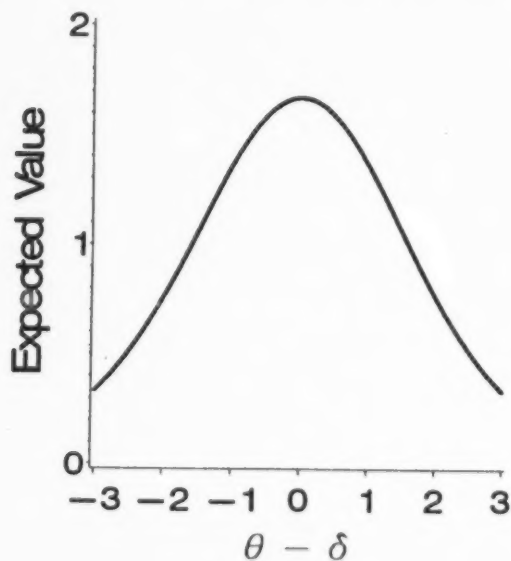
Tau Values: $-.5, -.25, 0, .25, \text{ and } .5$



Tau Values: $-1, -.5, 0, .5, \text{ and } 1$



Tau Values: $-2, -1, 0, 1, \text{ and } 2$



Tau Values: $-3, -1.5, 0, 1.5, \text{ and } 3$

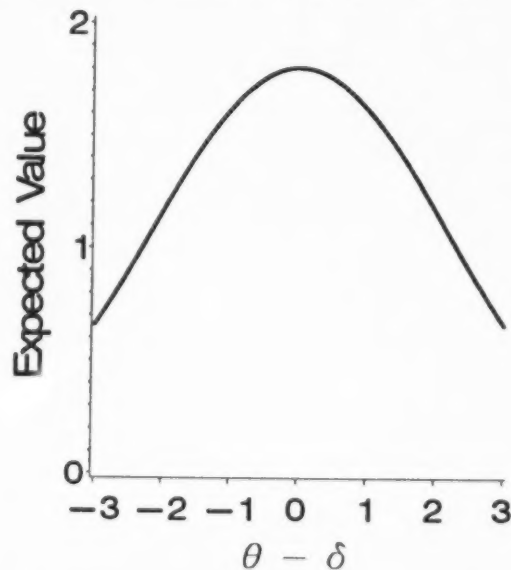


Figure 2. Response functions for a hypothetical 3-category item under the GGUM. Response functions are given for alternative values of the τ_a parameters while the value for α_i is held constant. The τ_a parameters for each function are ordered and equally spaced. The distance between successive τ_a parameters (i.e., interthreshold distance) is varied between response functions.

Figures 1 and 2 illustrate the response function of the GGUM for a hypothetical item with a 3-category response scale, where 0=disagree, 1=neither disagree nor agree (neutral), and 2=agree. In Figure 1, the discrimination parameter, α_i , varies from .5 to 30 while the subjective category thresholds, τ_{ik} , are held constant. As α_i increases, the response function becomes more peaked and takes on higher expected values. For extremely large values of α_i , the response function resembles a Guttman-like step function where successively higher order response categories are expected as the distance between θ_j and δ_i decreases. In Figure 2, α_i is held constant as the distance between equally spaced τ_{ik} values (i.e., the interthreshold distance) is varied. As the interthreshold distance grows from .25 to 1.5, the response function takes on larger expected values, but it becomes broader (i.e., more generalized). The role of α_i and τ_{ik} are, thus, quite distinctive within the GGUM.

To date, there has been only one empirical study of parameter recovery in the GGUM. Roberts, Donoghue and Laughlin (1998) have shown that when data conform perfectly to the GGUM and θ_j values are known to follow a normal distribution, then item parameters can be accurately estimated using a marginal maximum likelihood (MML) technique (Bock & Aitkin, 1981; Muraki, 1992) with samples of 750 or more respondents. Additionally, person locations (i.e., attitudes) can be accurately estimated with an expected a posteriori (EAP) technique (Bock & Mislevy, 1982) when there are approximately 15 to 20 items with 6 graded disagree-agree response categories per item. Thus, the minimum data requirements of the MML and EAP methods, as implemented in the GGUM, have been examined for the case where the data follow the model perfectly and the true distribution of θ_j is known.

Although these minimum data demands have been studied, questions about parameter estimability in the GGUM remain. For example, both the MML item parameter estimation and EAP person parameter estimation techniques require the specification of a prior distribution for θ_j . In their previous recovery study, Roberts et al. (1998) used a normal prior distribution that perfectly matched the true θ_j distribution. Thus, the consequences of using a normal prior distribution to estimate GGUM parameters when the true θ_j distribution is not normal are currently unknown. The normal distribution would often seem to be a reasonable choice for a prior distribution in the absence of information about the true θ_j distribution. In practice, however, the true distribution of θ_j will likely deviate from the normal distribution to at least some degree, and there is currently no information about the sensitivity of MML and EAP estimates to discrepancies between prior and true distributions of θ_j . If item and/or person parameters estimated in the GGUM are especially sensitive to the prior distribution assumption, then the utility of the MML and EAP estimation techniques will be compromised whenever information about the true distribution of θ_j is lacking. Although some information on sensitivity exists for binary cumulative models (Bock & Aitkin, 1981; Seong, 1990; Bartholomew, 1988), there is no such information about polytomous unfolding models. Consequently a major goal of this paper is to assess the effects of using a normal prior distribution for θ_j when the corresponding true θ_j distribution is not normal. A secondary goal is to assess the degree of improvement in parameter estimates that can be obtained by correctly specifying a nonnormal prior distribution that perfectly matches the corresponding true distribution of θ_j .

A related point can be made with regard to the numerical methods used to solve for GGUM parameter estimates in the MML and EAP algorithms. Each of these algorithms requires the evaluation of an integral taken over the prior distribution for θ_j . This evaluation is performed using numerical quadrature. However, there is currently no information about the sensitivity of parameter estimates in the GGUM to the number of quadrature points implemented in the MML and EAP estimation algorithms. Therefore, this paper will examine the effect that the number of quadrature points has on the estimates derived from the MML and EAP techniques.

The sensitivity of both MML estimation of item parameters and EAP estimation of person parameters to the correctness of the prior θ_j distribution and to the number of quadrature points utilized in the numerical integration process will be examined with parameter recovery simulations. In these simulations, the generating (true) and prior distributions for θ_j will be systematically varied along with the number of quadrature points used in the estimation algorithm. The role of the calibration sample size will also be investigated. These simulation results will provide much needed information about the robustness of MML and EAP parameter estimates for the GGUM.

Method

Design

The recovery simulations conducted in this study examined the effects of the following four variables:

1) type of generating (true) θ_j distribution

- normal
- bimodal-symmetric
- negatively skewed
- positively skewed

2) sample size

- N=500
- N=750
- N=1000
- N=2000

3) number of quadrature points

- 15 points
- 20 points
- 30 points

4) Correspondence between the prior and generating distributions

- matching prior distribution
- nonmatching normal prior distribution

These factors were partially crossed to produce a 4 (generating distribution) x 4 (sample size) x 3 (number of quadrature points) x 2 (correspondence between prior and generating distribution) design with one missing cell. The first two of these factors were between-replications factors whereas the last two were within-replications factors. The design was a partial factorial because it contained a missing cell which arose because it was logically impossible to have a normally distributed prior distribution that did not match the normally distributed generating distribution.

Generating Distributions

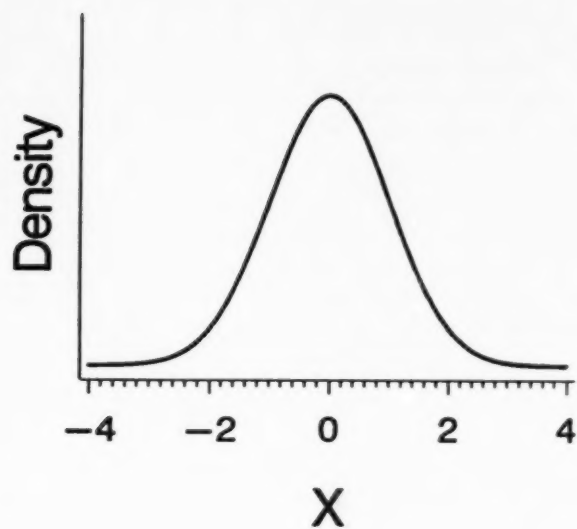
All of the generating distributions were standardized to have a mean of 0 and a variance of 1. The bimodal-symmetric distribution (henceforth referred to as the bimodal distribution) was developed by mixing two normal distributions with means of $\pm .894$ and variances of .2 where the mixing probabilities were equal to .5. The resulting distribution had a mean of 0, a variance of 1, a skew of 0 and a kurtosis of -1.28. The two skewed distributions were developed using formulas derived by Ramberg, Tadikamalla, Dudewicz, and Mykytka (1979). Each skewed distribution had a mean of 0, a variance of 1, a kurtosis of 1.6 and a skew of either plus or minus 1. Figure 3 illustrates the density functions for these generating distributions

Item Parameters and Response Generation

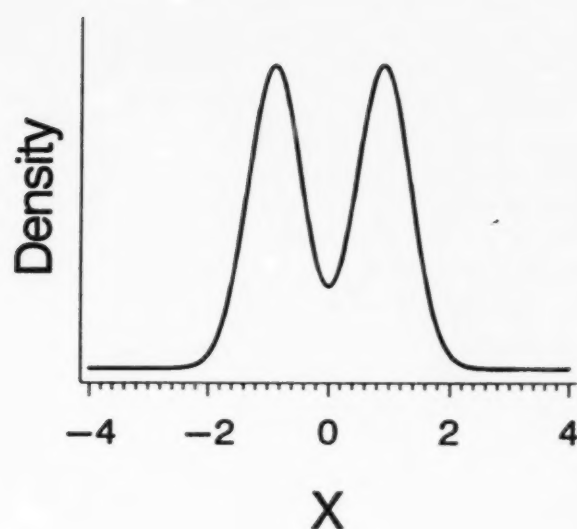
Twenty items with 6 response categories per item (0=strongly disagree, 1=disagree, 2=slightly disagree, 3=slightly agree, 4=agree, 5=strongly agree) were used in every condition. The 20 item locations were equally spaced between -2.0 and +2.0 on the latent continuum. Discrimination parameters for each item were randomly chosen from a uniform distribution spanning the interval of (.5, 2.0). Subjective category thresholds were determined in a sequential fashion. First, τ_{iC} was generated from a uniform (-1.4, -.4) distribution, and then successive values were derived from the recursive formula $\tau_{i,k-1} = \tau_{ik} - .25 + e_{i,k-1}$ for $k = C, \dots, 3, 2$, where $e_{i,k-1}$ denotes random error from a $N(0, .04)$ distribution. These item parameter values were consistent with past simulations of unfolding IRT models and were also consonant with analyses of real data.

The response of a simulee to an item was generated by computing the probability of observing a given response category as specified in equation 1. These response category probabilities were used to divide a (0,1) interval into discrete segments whose width corresponded to the relative magnitude of each response category probability. A random uniform deviate was then generated, and the segment of the (0,1) interval into which the random deviate fell determined the simulated observed response to that item.

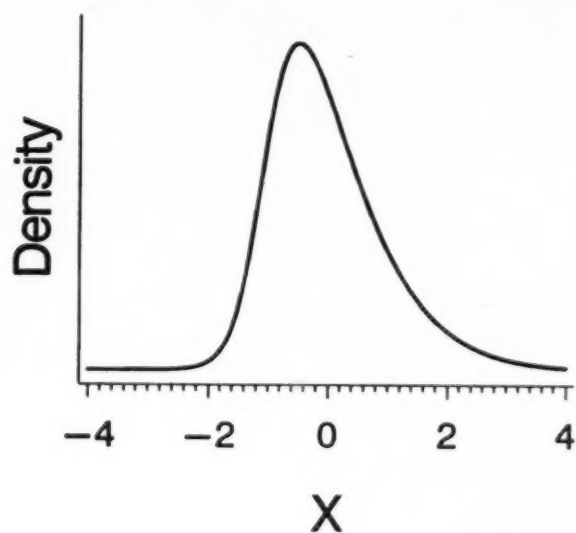
Normal Distribution



Bimodal Distribution



Positively Skewed Distribution



Negatively Skewed Distribution

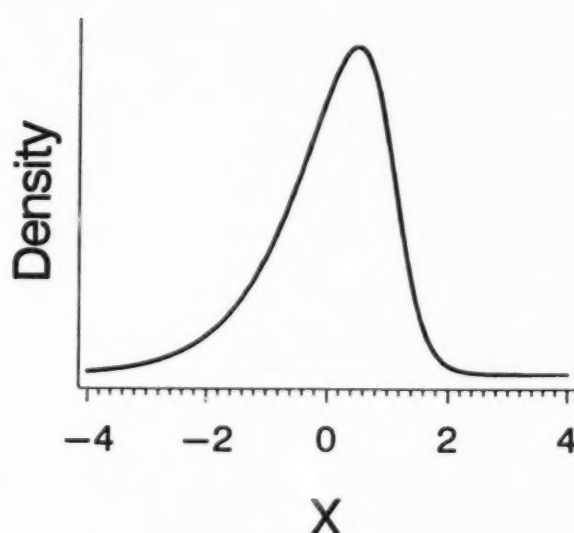


Figure 3. Probability density functions for the four θ_i generating distributions used in the simulation.

Parameter Estimation

Item parameters were estimated using a marginal maximum likelihood (MML) technique described in Roberts, Donoghue and Laughlin (1998). In this technique, θ_j parameters are numerically integrated out of the likelihood equation, and then the log of the marginal likelihood is maximized to find estimates of the item parameters. The iterative numerical integration-maximization process was continued until the change in any item parameter was less than .001. The quadrature points were always equally spaced along the $[-4, +4]$ interval regardless of the number of quadrature points or the type of prior distribution used in the integration process. On a given replication, item responses were generated, and then parameters were estimated under 6 different estimation conditions defined by the number of quadrature points (15, 20, 30) and whether the prior distribution matched the generating θ_j distribution (yes, no). The one exception to this process occurred when the generating distribution was normal, in which case, only a matching normal prior was investigated.

Once a given set of item parameters were estimated, the person parameters were then estimated using the expected a posteriori (EAP) method described in Roberts et al. (1998). This method uses the mean of the posterior θ_j distribution as the estimate for the j th individual given the observed responses, the estimated item parameters, and the prior distribution of θ_j .

The process of generating data and repeatedly solving for parameters was replicated 10 times in each θ_j generating distribution $\times$ sample size condition. Within each of these 10 replications, the generating item and person parameters were held constant, and new item responses were generated and subsequently analyzed.

Measures of Estimation Accuracy

Three measures of estimation accuracy were investigated in this study. The Root Mean Squared Error (RMSE) was the first of these measures. The RMSE provided an index of the average unsigned discrepancy between a set of true parameters and a corresponding set of estimates. The RMSE was calculated across all the parameters of a given type in any single replication. For example, the RMSE of item location estimates from a particular replication was computed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^I (\hat{\delta}_i - \delta_i)^2}{I}}, \quad (2)$$

where:

- δ_i = the true location of the i th item on the attitude continuum,
- $\hat{\delta}_i$ = the estimated location for the i th item on the attitude continuum, and
- I = the number of items on the test (i.e., 20).

Analogous quantities were computed for the item discrimination, subjective response category threshold and person parameters in a given replication.

The Pearson correlation between estimated and true parameters was the second measure of accuracy utilized in the investigation. This correlation was computed across a given set of parameters (i.e., α_i , δ_i , τ_{ik} , or θ_j) within each replication. Therefore, it provided a simple index of the degree of linearity between estimated and true parameter values. We refer to this correlation as a recovery correlation and denote it as RCORR in subsequent sections of this report.

The third and final measure of accuracy was the average absolute discrepancy between estimated and true response functions on the θ interval of $[-3, 3]$. This measure was calculated by integrating the absolute difference between the response function of the i th item computed with true parameters and that computed with estimated parameters, and then dividing the result by 6 (i.e., the length of the evaluation interval):

$$D_i = \int_{-3}^3 \sum_{z=0}^C z \left| \hat{Pr}[Z_i = z | \theta] - Pr[Z_i = z | \theta] \right| d\theta. \quad (3)$$

Note that $\hat{Pr}[Z_i = z | \theta]$ and $Pr[Z_i = z | \theta]$ represent the value of Equation 1 when calculated from estimated and true item parameters, respectively. The integral in equation 3 was evaluated numerically for each item using a globally adaptive scheme based on Gauss-Kronrod rules (Piessens, deDoncker-Kapenga, Uberhuber, Kahaner, 1983), and then the resulting D_i values were averaged across all the items within a given replication. The resulting measure, denoted as *AAD*, provided an index of the average similarity between the estimated response functions and the true functions for a given set of items. Some researchers have argued that an index like *AAD* is the most useful measure of overall estimation accuracy because item parameter estimates can be substantially different from true values yet still yield estimated response functions that are quite similar to their corresponding true functions (Hulin, Lissak & Drasgow, 1982; Linn, Levine, Hastings & Wardrop, 1981).

Analysis of Accuracy Measures

Each accuracy measure was analyzed using two different ANOVAs in an effort to provide meaningful and unique results in the presence of a missing cell within the factorial design. The first analysis was a univariate split-plot ANOVA conducted using only those data from the normal prior distribution conditions. This analysis examined all main effects and interactions involving the generating distribution for θ_j , the sample size, and the number of quadrature points used. The primary purpose of this analysis was to determine the impact of using a normal prior distribution in the parameter estimation algorithm when the true θ_j distribution was not normal - a situation which is likely to occur in practice to at least some degree. The effect of the number of quadrature points used and the effect of sample size were also of interest here, as were the interrelationships of these variables with the sensitivity to the normal prior distribution.

The second analysis was a univariate split-split-plot ANOVA and utilized all the data except those from the normal generating distribution conditions. In this second analysis, the main effects and interactions involving the generating distribution for θ_j , the sample size, the number of quadrature points, and the correspondence between the prior and generating distributions were all tested. The primary focus of this second analysis was to assess the improvement in parameter estimation accuracy obtained from perfectly matching a nonnormal prior distribution with a corresponding nonnormal generating distribution for θ_j , relative to the estimation accuracy encountered with a normal prior distribution. Again, the effect of the number of quadrature points, the effect of sample size and the interactive effects of these variables on the impact of the prior distribution was also of interest.

When considering both the first and second analyses, there were a total of 18 dependent measures studied for those effects not involving the match between the prior and generating distribution. Therefore, the Type I error rate was set to $\alpha = .05/18 = .00278$ when testing those effects. In contrast, only 9 dependent measures were studied with regard to those effects involving the matching factor, and thus, the Type I error rate was held at $\alpha = .05/9 = .0056$ for these effects. The proportion of variance associated with each effect (η^2) was calculated for each dependent measure. An effect was deemed to be worthy of interpretation if it was both statistically significant and accounted for at least 5% of the variation in a given dependent measure.

Results

MML Estimates of Item Parameters

Analysis 1. Table 1 displays the proportion of variance, η^2 , associated with each of the ANOVA effects in the first analysis of the RMSE, RCORR, and AAD measures. Recall that this first analysis was limited to those replications where a normal prior distribution for θ_j was used with either a normal, bimodal, positively skewed or negatively skewed generating distribution. As shown in Table 1, the RMSE measures for δ_i , α_i , and τ_{ik} were primarily a function of the generating distribution for θ_j , the sample size, and the number of quadrature points used in the numerical integration process. Figure 4 displays the average RMSE values obtained for each type of item parameter under each θ_j generating condition. Interestingly, the RMSE values were statistically similar across the generating conditions with the exception of the bimodal distribution condition in which the highest RMSE was incurred. Posthoc comparison of all pairwise means with Tukey's HSD test confirmed this fact. Nonetheless, the absolute degree of error associated with any of the generating distributions was both similar and reasonable. In the bimodal distribution condition, the average RMSE represented 12.4%, 32.7% and 38.8% of the standard deviations of true δ_i , α_i , and τ_{ik} parameters, respectively. The corresponding quantities for the normal distribution condition were 9.3%, 26.8% and 32.0%. The magnitude of these error indices also suggested that δ_i parameters were more easily estimated than either α_i or τ_{ik} parameters.

Table 1. η^2 values for ANOVA effects from the first analysis of accuracy measures.^a

Effect	Analysis 1 RMSE				Analysis 1 RCORR				AAD
	$\hat{\theta}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\theta}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\tau}$	
G	.23	.11	.06	.08	.20	.19	.09	.07	.06
N	.08	.20	.15	.48	.08	.25	.44	.64	.13
G x N	.29	.04	.01	.02	.32	.04	.03	.01	.01
Q	.05	.20	.47	.15	.01	.00	.00	.00	.52
G x Q	.00	.01	.02	.00	.00	.00	.00	.00	.06
N x Q	.00	.00	.02	.00	.00	.00	.00	.00	.01
G x N x Q	.00	.01	.00	.00	.00	.00	.00	.00	.00

^a G=type of generating distribution for θ , N=sample size, Q=number of quadrature points. Statistically significant effects are denoted by boldface type. Significance was determined by a univariate, split-plot analysis of variance with $\alpha=.00278$. Effects were deemed to warrant interpretation when they were both statistically significant and had corresponding η^2 values greater than or equal to .05.

Figure 5 displays the average RMSE incurred for δ_i , α_i , and τ_{ik} in the first analysis as a function of calibration sample size. As expected, the amount of error associated with item parameter estimates decreased as the sample size increased from 500 to 2000. However, the increased precision afforded by larger samples began to wane after the sample size reached 1000. The η^2 values in Table 1 suggest that the sample size effect was most pronounced for the estimation of τ_{ik} parameters. This was corroborated when the average RMSE indices were expressed as a percentage of the standard deviation of the corresponding true parameters. The percentages were equal to 12.5% for δ_i , 34.7% for α_i , and 43.2% for τ_{ik} in the N=500 condition. Those for the N=2000 condition were equal to 7.5%, 25.2% and 25.4% respectively. Thus, the largest reduction in this error index occurred with the τ_{ik} parameters.

Figure 6 illustrates the average RMSE for δ_i , α_i , and τ_{ik} as a function of the number of quadrature points. For each item parameter, there was a very noticeable decline in RMSE as the number of quadrature points increased from 15 to 20, after which there was little gain in precision associated with more quadrature points. The η^2 values in Table 1 suggest that the number of quadrature points had the most impact on the RMSE for α_i followed by that for δ_i . This impact on α_i estimates was corroborated when the RMSE was expressed as a percentage of the standard deviation of true parameter values. With 15 quadrature points, these percentages were equal to 12.7%, 38.6%, and 39.6% for δ_i , α_i , and τ_{ik} , respectively. The corresponding percentages for the 20 quadrature point condition were 9.0%, 25.9% and 32.3%.

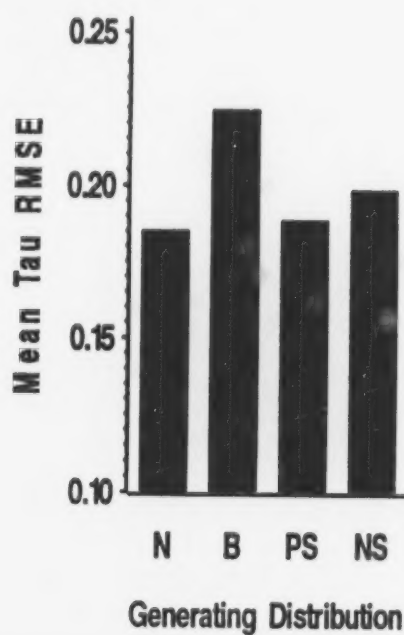
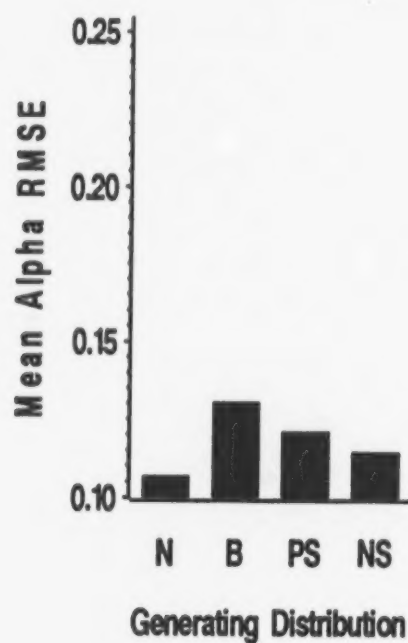
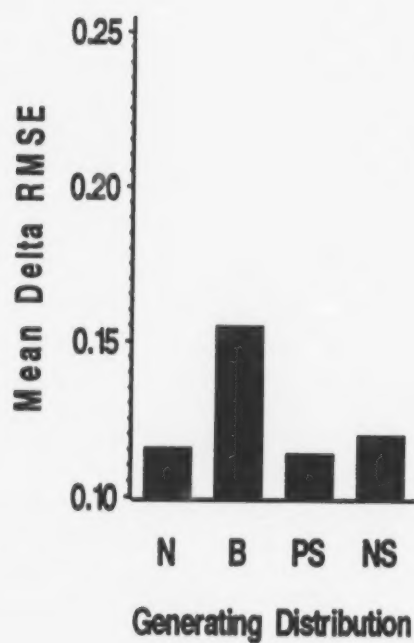


Figure 4. Mean RMSE indices for the item parameter estimates derived from different generating distributions for θ_j . N=normal distribution, B=bimodal distribution, PS=positively skewed distribution, and NS=negatively skewed distribution. Results are from Analysis 1, which included only those data from conditions with a normal prior distribution for θ_j .

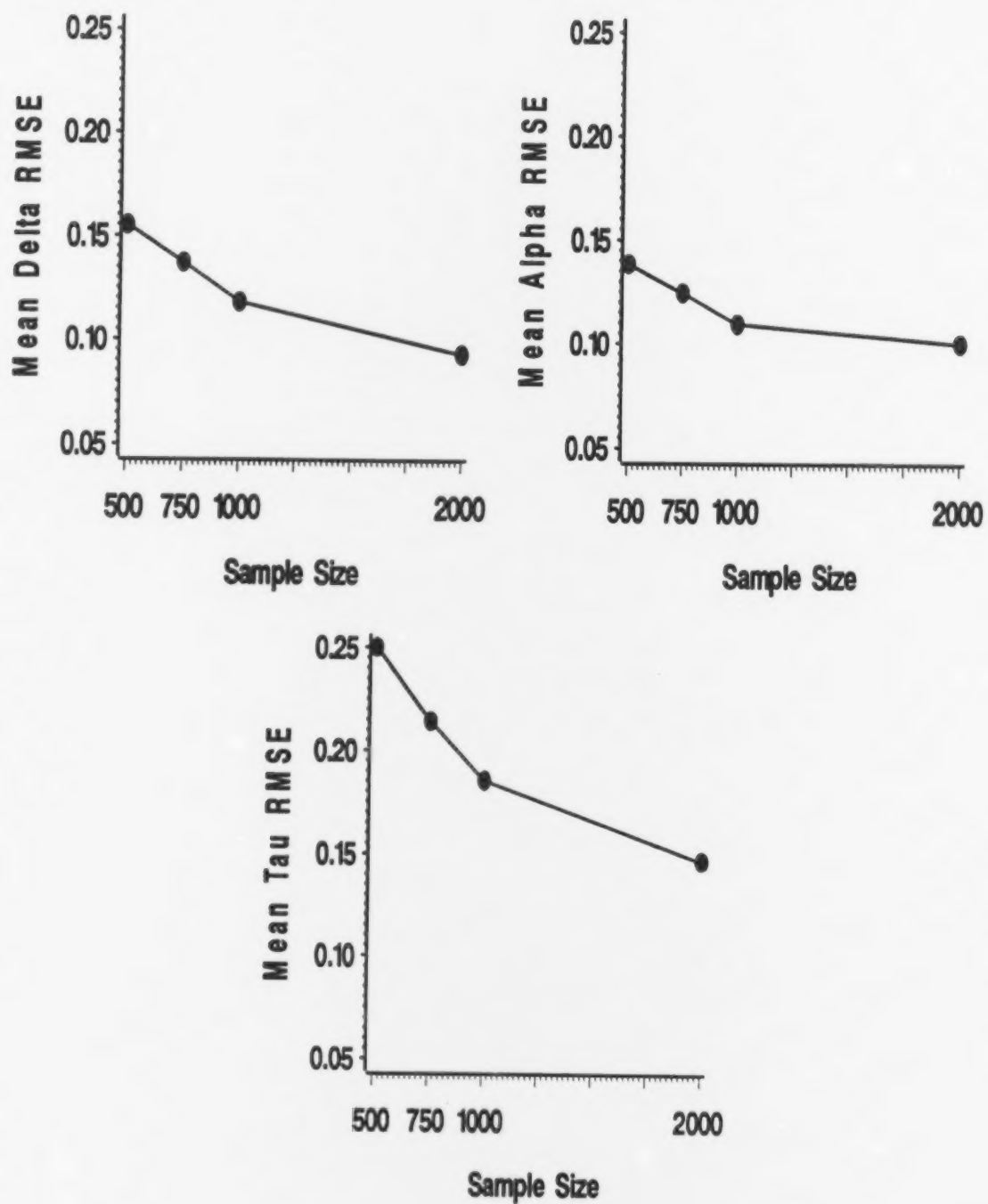


Figure 5. Mean RMSE indices for the item parameter estimates derived under different sample size conditions. Results are from Analysis 1, which included only those data from conditions with a normal prior distribution for θ_j .

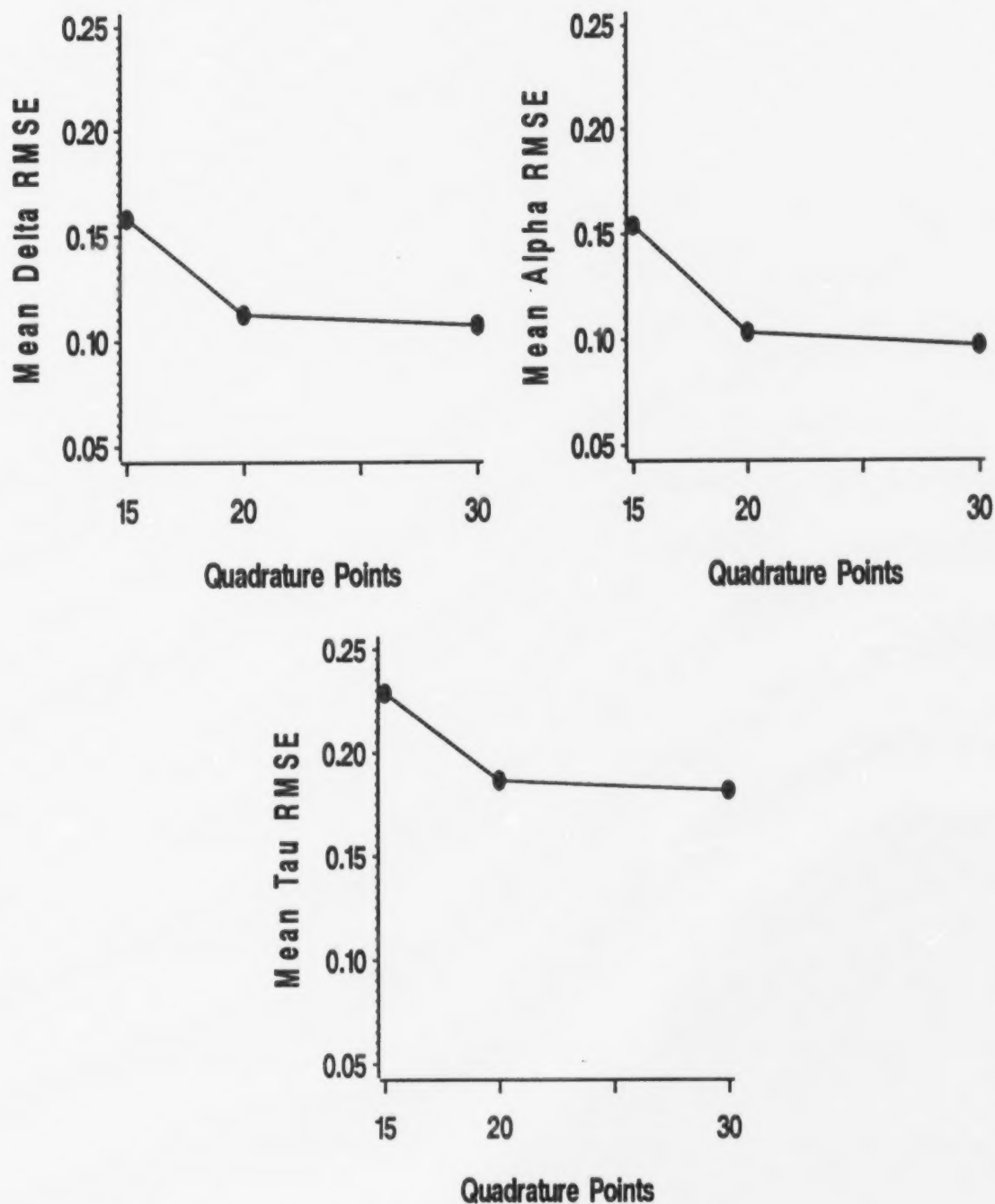


Figure 6. Mean RMSE indices for the item parameter estimates derived using different numbers of quadrature points in the MML procedure. Results are from Analysis 1, which included only those data from conditions with a normal prior distribution for θ_j .

The average RCORR for δ_i , α_i , and τ_{ik} was primarily a function of the generating distribution for θ_j and the sample size. Figure 7 illustrates the average RCORR values for δ_i , α_i , and τ_{ik} as a function of the generating distribution. Departures from linearity between true and estimated parameters were generally greatest in the bimodal θ_j generating condition, although the average RCORR index in that condition was still high for all three item parameters (i.e., .94 or greater). Figure 8 shows the average RCORR values for δ_i , α_i , and τ_{ik} as a function of sample size. The average RCORR values consistently increased as the sample size increased. However, there was little change in average RCORR values for δ_i and α_i after the sample size reached 1000, whereas the RCORR values for τ_{ik} continued to improve in a relative sense.

The average AAD values were primarily a function of the generating distribution for θ_j , the sample size, the number of quadrature points, and the generating distribution $\times$ quadrature points interaction. The η^2 values from Table 1 indicate that the number of quadrature points was the most pronounced of these effects. Figure 9 shows the average AAD values corresponding to each effect. The average AAD values shown in Figure 9 were consistently small in absolute magnitude (e.g., less than .18 on a 6 point response scale). With regard to the generating distribution for θ_j , the largest mean AAD was incurred with the skewed distributions (.114 and .112 for the positively skewed and negatively skewed distributions, respectively) although the maximum difference in AAD between any two distribution conditions was small (e.g., .025). As one might expect, the average AAD decreased as a function of sample size falling from .125 to .082 as the sample size increased from 500 to 2000. The average AAD also fell as the number of quadrature points increased. However, decreases in AAD were minor beyond 20 quadrature points, and this pattern was consistently observed for all θ_j generating distributions. The generating distribution $\times$ quadrature point interaction appeared primarily due to the unusually low level of AAD for the bimodal condition when 15 quadrature points were used.

Analysis 2. Table 2 gives the η^2 values for the ANOVA effects in Analysis 2. Recall that in this analysis, only the non-normal θ_j generating conditions were considered and the effects of both matching and non-matching (normal) prior distributions were examined. As in the previous analysis, the main effects of sample size and number of quadrature points were evident when considering the average RMSE for δ_i , α_i , and τ_{ik} . The pattern of these effects were very similar to those found in Analysis 1 (see Figures 5 and 6) in that increasing sample size decreased the amount of error in the estimates, as did increasing the number of quadrature points. Again, the decreases in estimation error were minimal with more than 1000 subjects or more than 20 quadrature points, respectively. Interestingly, with regard to average RMSE, there was no effect involving the match between the generating and prior θ_j distributions which met the criteria for interpretation. This suggested that the effects of matching a nonnormal true θ_j distribution with a correctly specified prior distribution were small relative to those mentioned above.

The main effect of sample size appeared to substantially influence the average RCORR values for δ_i , α_i , and τ_{ik} . As the sample size increased, the correlation between true and estimated

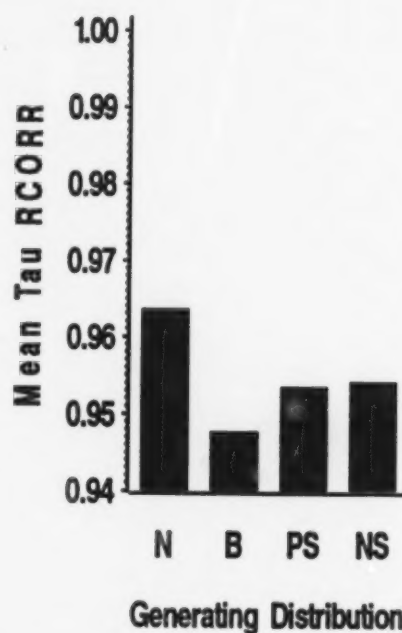
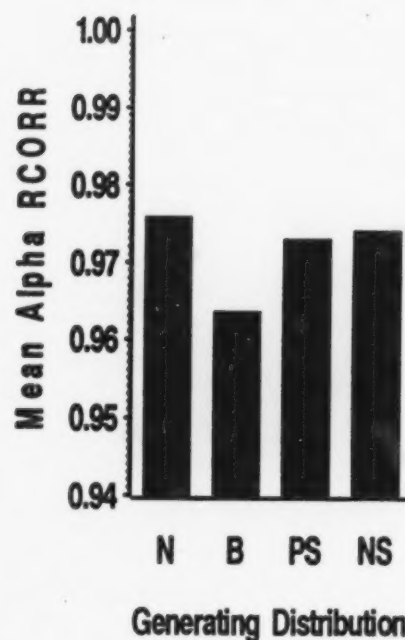
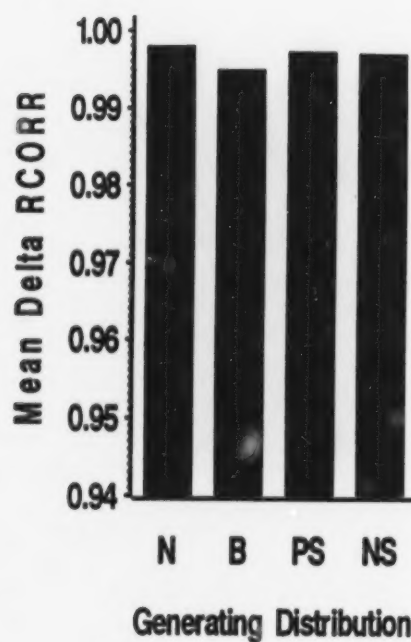


Figure 7. Mean RCORR indices for the item parameter estimates derived from different generating distributions for θ_j . N=normal distribution, B=bimodal distribution, PS=positively skewed distribution, and NS=negatively skewed distribution. Results are from Analysis 1, which included only those data from conditions with a normal prior distribution for θ_j .

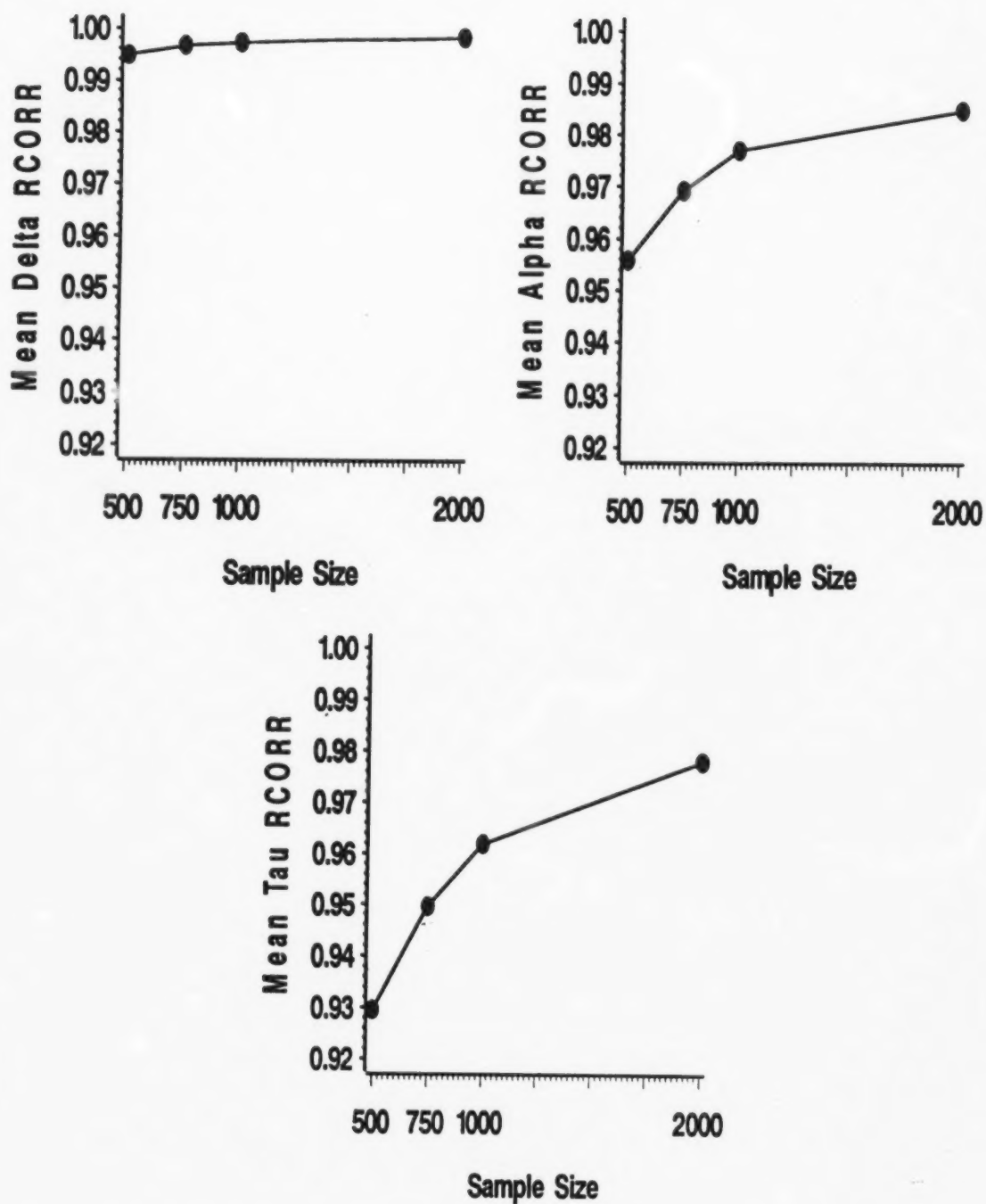


Figure 8. Mean RCORR indices for the item parameter estimates derived under different sample size conditions. Results are from Analysis 1, which included only those data from conditions with a normal prior distribution for θ_j .

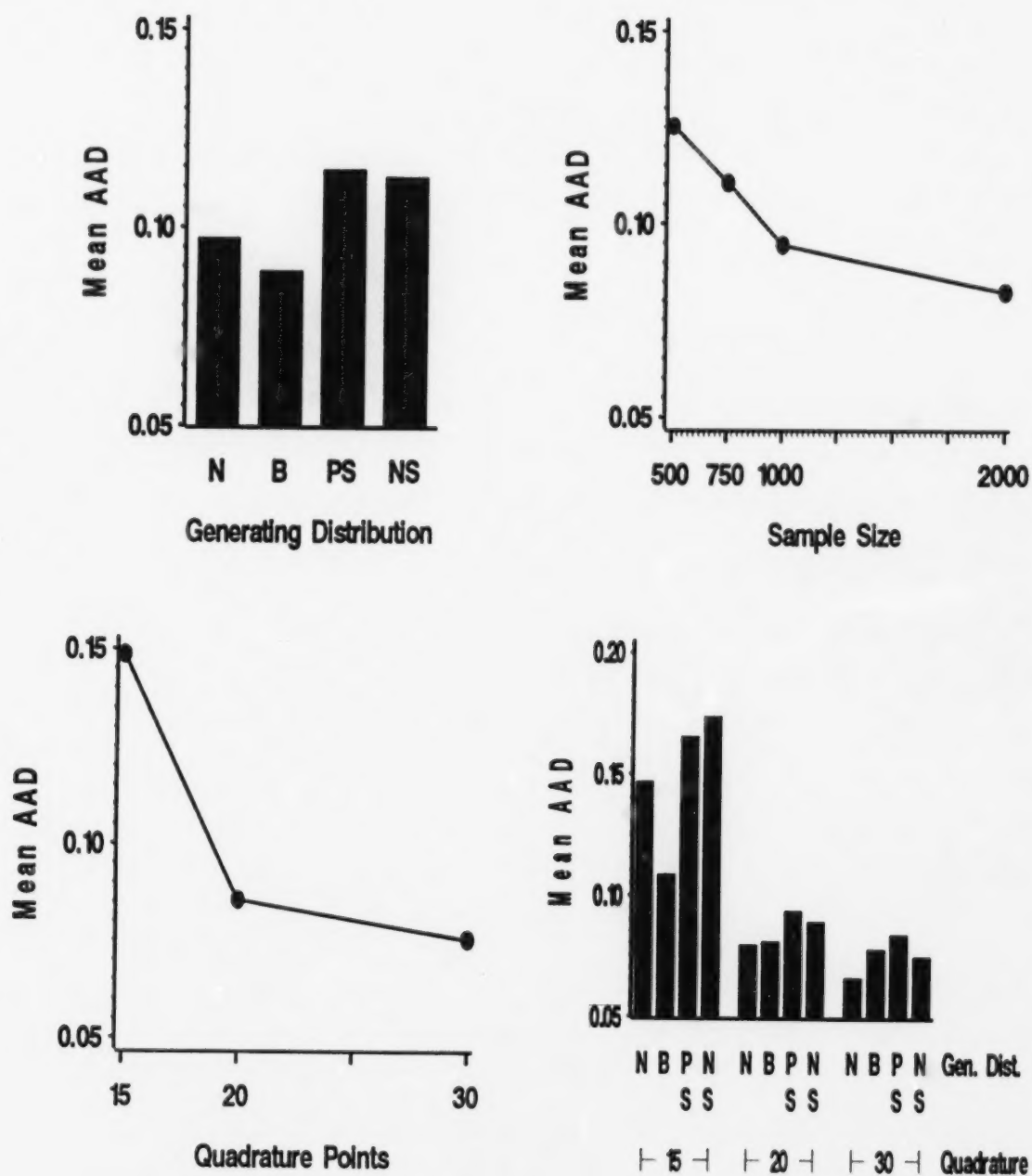


Figure 9. Mean AAD indices for the item parameter estimates derived from different generating distributions for θ_j , different sample size conditions, different quadrature point conditions, and different generating distribution x quadrature point conditions. N=normal distribution, B=bimodal distribution, PS=positively skewed distribution, and NS=negatively skewed distribution. Results are from Analysis 1, which included only those data from conditions with a normal prior distribution for θ_j .

Table 2. η^2 values for ANOVA effects from the second analysis of accuracy measures.^a

Effect	Analysis 2 RMSE				Analysis 2 RCORR				AAD
	$\hat{\theta}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\tau}$	$\hat{\theta}$	$\hat{\delta}$	$\hat{\alpha}$	$\hat{\tau}$	
G	.10	.04	.00	.02	.08	.09	.03	.01	.01
N	.07	.18	.18	.47	.06	.28	.47	.65	.12
G x N	.13	.01	.02	.01	.15	.04	.04	.01	.01
Q	.07	.20	.42	.14	.01	.00	.00	.00	.50
G x Q	.00	.01	.03	.00	.00	.00	.00	.00	.03
N x Q	.00	.00	.01	.00	.00	.00	.00	.00	.00
G x N x Q	.00	.00	.00	.00	.00	.00	.00	.00	.00
M	.11	.01	.03	.03	.09	.01	.02	.02	.05
G x M	.06	.04	.02	.02	.07	.02	.02	.00	.04
N x M	.03	.00	.00	.00	.04	.01	.00	.00	.00
G x N x M	.11	.00	.00	.00	.13	.00	.00	.00	.00
Q x M	.00	.00	.00	.00	.00	.00	.00	.00	.00
G x Q x M	.00	.01	.00	.00	.00	.00	.00	.00	.01
N x Q x M	.00	.00	.00	.00	.00	.00	.00	.00	.00
G x N x M x Q	.00	.00	.00	.00	.00	.00	.00	.00	.00

^a G=type of generating distribution for θ_j , N=sample size, Q=number of quadrature points, M=match (correspondence) between generating and prior distributions for θ_j . Statistically significant effects are denoted by boldface type. Significance was determined by a univariate, split-split-plot analysis of variance with $\alpha=.00278$ for those effects not involving the matching factor and $\alpha=.00556$ for those effects that did involve the matching factor. Effects were deemed to warrant interpretation when they were both statistically significant and had corresponding η^2 values greater than or equal to .05.

parameters consistently increased. As the sample size varied from 500 to 2000, the average RCORR values increased from .995 to .998 for δ_i , .956 to .986 for α_i , and .930 to .978 for τ_{ik} . There was also a small main effect of generating distribution for the δ_i parameters. Specifically, the average RCORR value was slightly lower in the bimodal generating distribution condition (.995) than for either of the skewed distribution conditions (.997). No effect involving the match between generating and prior θ_j distributions was deemed worthy of interpretation when considering the RCORR index.

The average AAD values in the second analysis were primarily a function of the sample size, the number of quadrature points and the correspondence between the prior distribution for θ_j and the generating distribution. Moreover, the number of quadrature points had the largest effect as indicated by the η^2 values in Table 2. The average AAD values consistently decreased with

increasing sample size in a fashion analogous to that seen in the first analysis (see Figure 9). The mean AAD decreased from .136 to .095 as the sample size increased from 500 to 2000. Similarly, the effect of the number of quadrature points was analogous to that from the first analysis (see Figure 9) and suggested that estimation error decreased as the number of quadrature points increased, although little error reduction was achieved with more than 20 quadrature points. The average AAD values were equal to .16, .10, and .09 for the 15, 20 and 30 quadrature point conditions, respectively. Finally, the effect of the consistency between the generating and prior θ_j distributions was in an intuitive direction. Specifically, the mean AAD was equal to .126 in those conditions in which the normal prior distribution did not match the generating distribution, whereas it was equal to .105 when the prior distribution matched the generating distribution. As seen in Table 2, this effect accounted for only 5% of the variation in AAD scores, and the absolute magnitude of this difference was small in a practical sense.

The most striking feature of the second analysis was that, with the exception of mean AAD scores, the other measures of item parameter recovery accuracy did not substantially depend on any ANOVA effects involving the match between the prior and generating distributions for θ_j . This suggested that item parameter recovery in the GGUM using the MML algorithm was fairly robust to the departures between the prior and generating distributions examined here.

EAP Estimates of Person Parameters

Analysis 1. Table 1 gives the η^2 indices associated with the accuracy measures for EAP estimates of the θ_j parameters in the first analysis in which the prior distribution was always the normal distribution. The results from this analysis indicated that the average RMSE for θ_j parameters was primarily a function of the generating distribution for θ_j and the interaction of sample size and generating distribution. However, the main effects of sample size and number of quadrature points also met the criteria for interpretation. As expected, the RMSE decreased as the number of quadrature points increased up to 20 points, after which only small decreases in estimation error were observed. Specifically, the average RMSE was equal to .245, .211 and .204 for the 15, 20 and 30 point conditions, respectively. The other effects on average RMSE are shown in Figure 10. The main effect of generating distribution was due to a large average RMSE value for the positively skewed generating condition relative to the other generating conditions. Similarly, the main effect of sample size was due to an unusually large amount of average RMSE found in the $N=750$ condition. The interaction of these two factors was primarily due to the fact that the negatively skewed generating condition was associated with the highest RMSE when the sample size was limited to 500. However, with larger sample sizes, the positively skewed condition exhibited the largest amounts of error, especially when $N=750$.

Results for the average RCORR values were similar to those for the RMSE index in that there were interpretable effects associated with the generating distribution for θ_j , the sample size, and the interaction between these two factors. The main effect of generating distribution indicated that the linear relationship between true and estimated θ_j was significantly smaller for the positively skewed generating condition, although all the average RCORR values were above

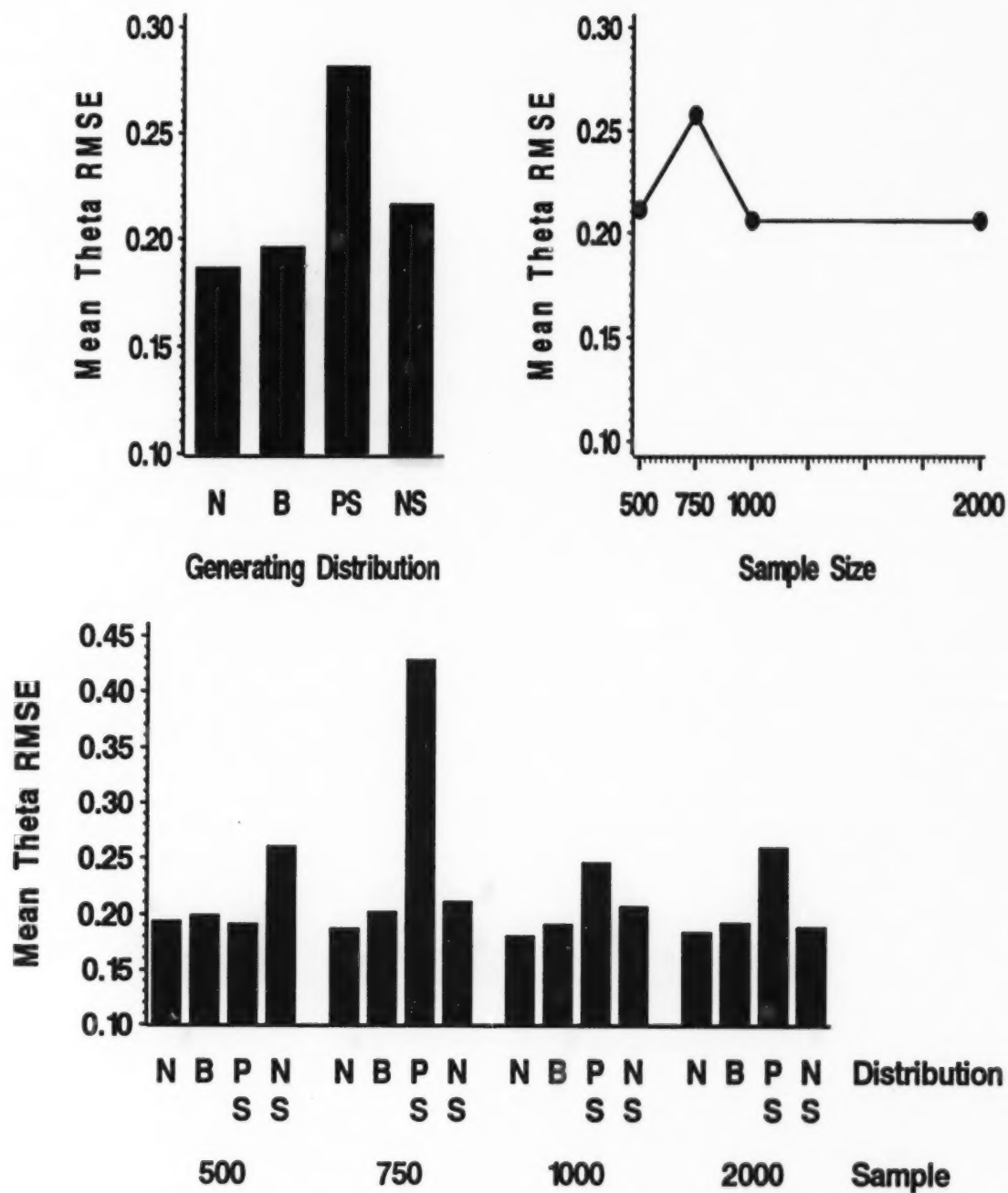


Figure 10. Mean RMSE indices for the person parameter estimates derived from different generating distributions for θ_j , different sample size conditions, and different generating distribution x sample size conditions. N=normal distribution, B=bimodal distribution, PS=positively skewed distribution, and NS=negatively skewed distribution. Results are from Analysis 1, which included only those data from conditions with a normal prior distribution for θ_j .

.95 in all conditions. The sample size main effect was due to a decreased linear relationship between true and estimated parameters when the sample size was equal to $N=750$, in which case, the average RCORR equaled .963. All other sample size conditions produced larger, relatively similar RCORR values (e.g., .979 on average). Finally, the generating distribution $\times$ sample size interaction revealed smaller average RCORR values for the positively skewed generating distribution, but only when the sample size was greater than 500. For the $N=500$ condition, the negatively skewed generating condition produced the smallest average RCORR value. Thus, the results from the analysis of RCORR values were consistent with those presented earlier for the RMSE index.

Analysis 2. Table 2 shows the η^2 values for the second analysis of the recovery accuracy measures for θ_j . Recall that the second analysis included the effects of correspondence between the generating and prior distributions. With regard to the RMSE index, there were interpretable effects due to the number of quadrature points, the generating distribution for θ_j , the sample size, and the interaction of these latter two factors. The patterns associated with these effects were highly similar to those from the first analysis (see Figure 10), and thus, they will not be discussed further. More importantly, the average RMSE for θ_j was slightly lower when there was correspondence between generating and prior distributions (.188) as opposed to when these distributions did not match (.231). Additionally, there was an interaction between generating distribution and correspondence between generating and prior distributions which is displayed in the top panel of Figure 11. When the generating and prior distributions did not match, the error found with the positively skewed generating distribution was disproportionately large. However, the differences in average RMSE values attenuated between generating conditions when the prior matched the generating distribution. There was also a 3-way interaction involving generating distribution $\times$ correspondence between prior and generating distributions $\times$ sample size. This interaction is illustrated in the lower panel of Figure 11. The interaction was due to the fact that the increased error encountered when using a nonmatching prior with a positively skewed distribution was disproportionately large for the $N=750$ condition while it was almost absent in the $N=500$ condition. In this latter condition, the most error was encountered when the nonmatching prior was used with the negatively skewed distribution.

With regard to the average RCORR values, there were interpretable effects due to the generating distribution for θ_j , the sample size and the interaction of these two factors. Each of these effects was very similar to that reported in the first analysis and will not be discussed further. Additionally, there were several effects that involved the correspondence between prior and generating distributions for θ_j , and these effects were consistent with those found for the RMSE index. Specifically, the average correlation between true and estimated θ_j was higher when the distributions matched (e.g., .984 versus .971). There was also a 2-way interaction involving the generating distribution $\times$ correspondence between prior and generating distributions. The mean RCORR values associated with this interaction are shown in the upper panel of Figure 12. The interaction was due to the fact that average RCORR values were disproportionately

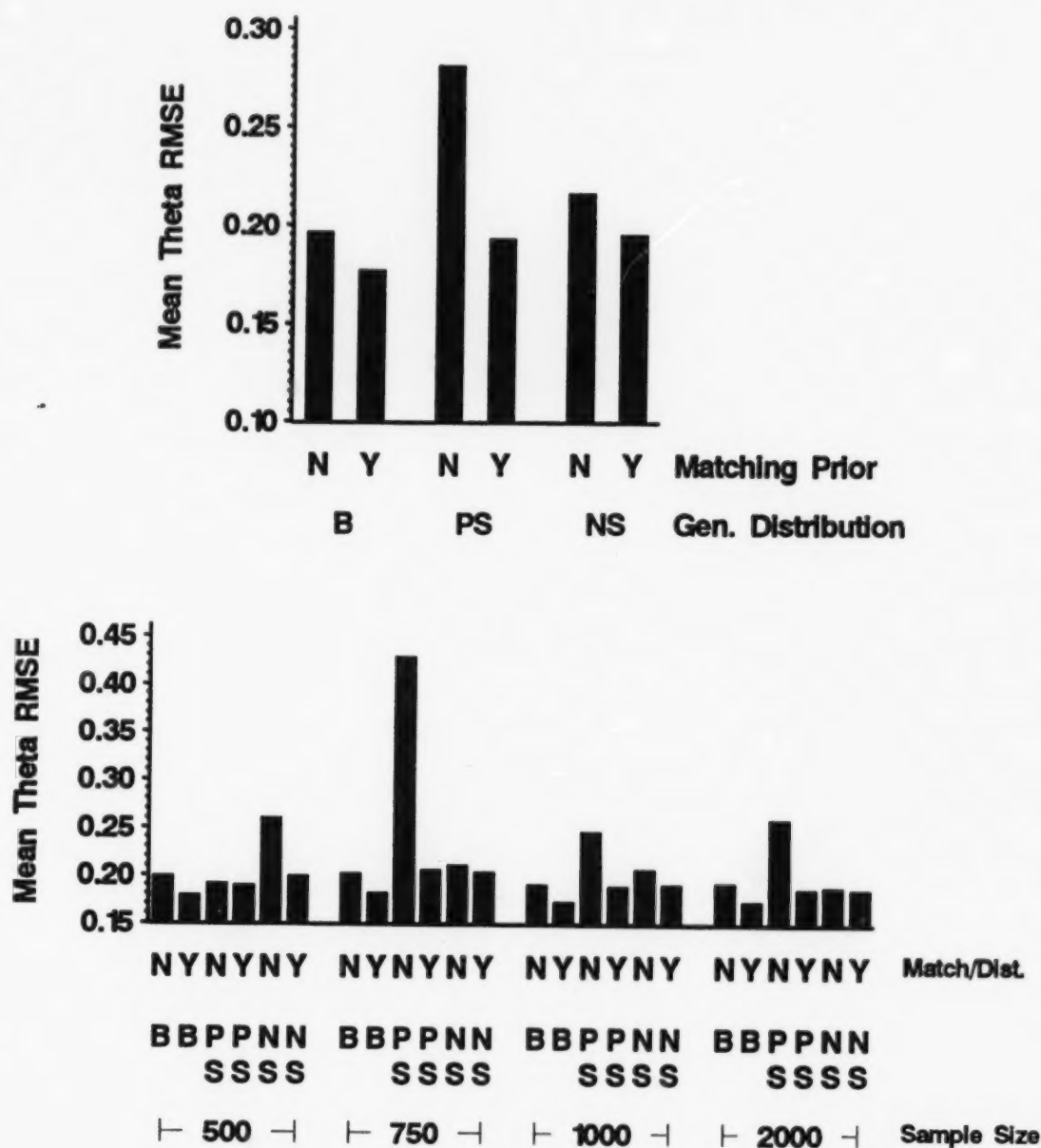


Figure 11. Mean RMSE indices for the person parameter estimates derived from different generating distributions for θ_j with either matching or nonmatching (normal) prior distributions (top panel). These same results are illustrated further by sample size (bottom panel). B=bimodal distribution, PS=positively skewed distribution, and NS=negatively skewed distribution, Y=matching prior distribution, N=nonmatching (normal) prior distribution. Results are from Analysis 2, which included only those data from conditions with a nonnormal generating distribution for θ_j and a matching or nonmatching prior distribution for θ_j .

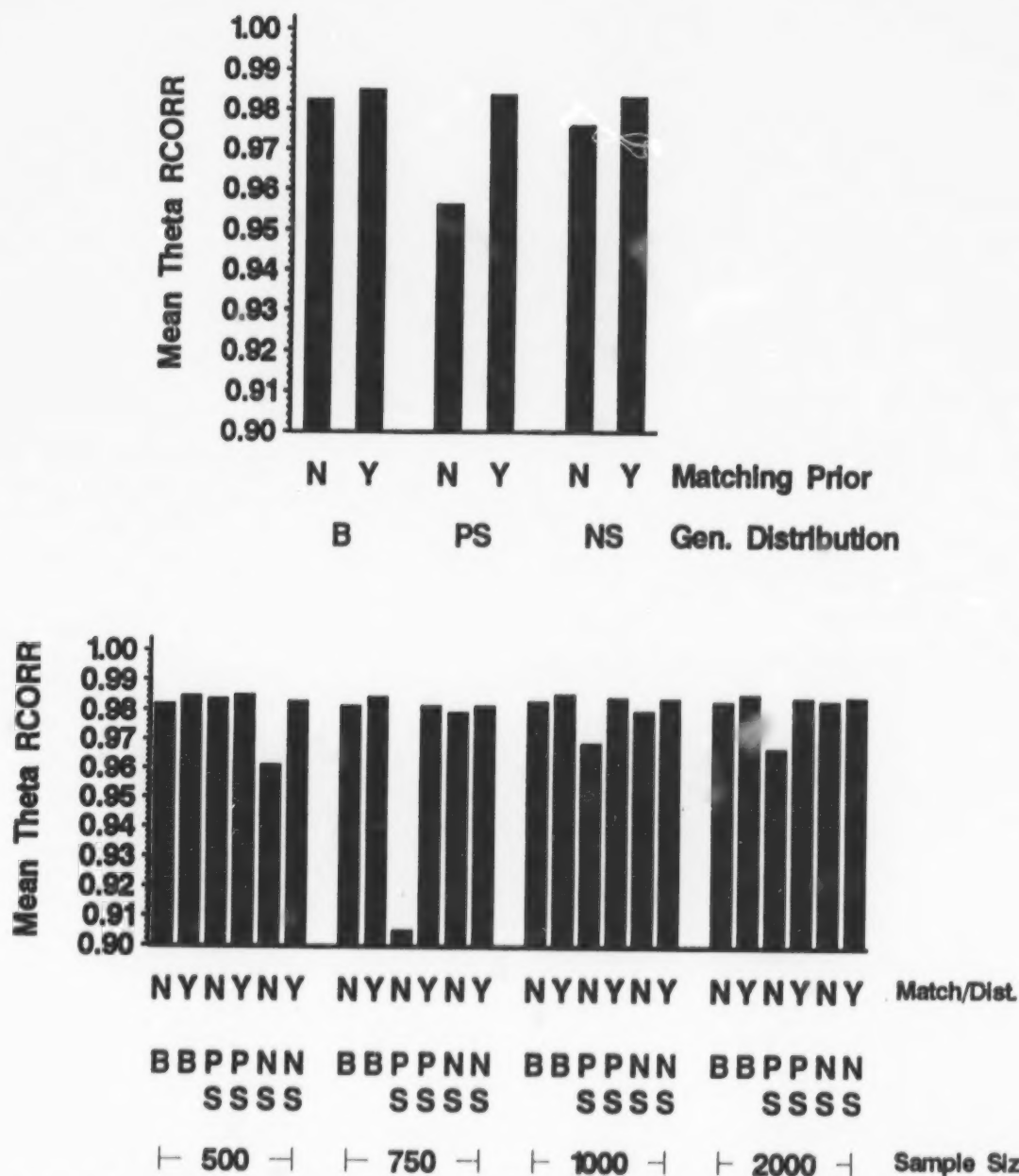


Figure 12. Mean RCORR indices for the person parameter estimates derived from different generating distributions for θ_j with either matching or nonmatching (normal) prior distributions (top panel). These same results are illustrated further by sample size (bottom panel). B=bimodal distribution, PS=positively skewed distribution, and NS=negatively skewed distribution, Y=matching prior distribution, N=nonmatching (normal) prior distribution. Results are from Analysis 2, which included only those data from conditions with a nonnormal generating distribution for θ_j , and a matching or nonmatching prior distribution for θ_j .

smaller for the positively skewed generating condition when the prior distribution did not match the generating distribution. There was also a 3-way interaction involving generating distribution x correspondence between prior and generating distributions x sample size. The average RCORR values describing this interaction are given in the lower panel of Figure 12. The interaction occurred because the positively skewed distribution produced disproportionately lower RCORR values when the prior did not match the generating distribution, but this occurred only for sample sizes greater than 500. Moreover, this effect was unexpectedly large when the sample size was equal to 750. In contrast, the negatively skewed distribution produced the most obvious reduction in RCORR values when the prior did not match and the sample size was equal to 500.

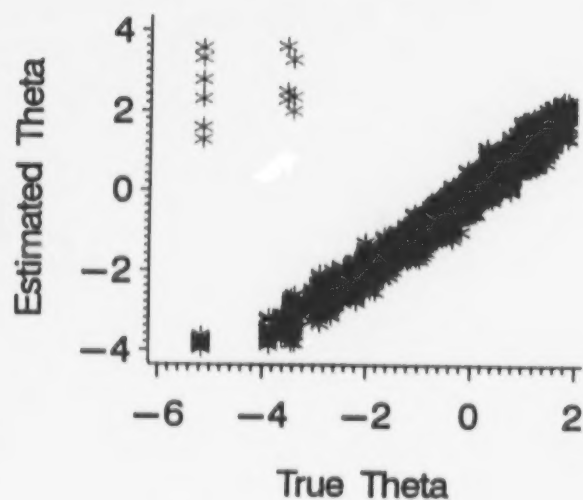
Post-hoc Analyses of θ_j Estimates

As mentioned above, the magnitude of error in person estimates was sometimes unusually high when a nonmatching normal prior distribution was used with a skewed generating (true) distribution. Moreover, this result interacted with sample size in a unintuitive manner. To understand these results better, the relationship between true and estimated θ_j values was plotted for the condition in which the worst recovery accuracy was encountered - namely, the condition in which a positively skewed generating distribution was used with a nonmatching normal prior and a sample size of $N=750$. The scatterplot for this condition is shown in the upper right panel of Figure 13. Note that data from all the replications in this condition are shown in the scatterplot. Thus, there are 10 points per simulee and each of these 10 points correspond to the same true θ_j value. It is obvious from the scatterplot that estimation error was exceedingly high for simulees in the right tail (i.e., the skewed tail) of the underlying θ_j distribution. Further examination of these simulees revealed that they had extremely low item scores indicative of strong disagreement with all items. In essence, these simulees strongly disagreed with most, if not all, items because no items were close to their extreme θ_j positions. The remaining panels in Figure 13 show analogous scatterplots for other conditions in which the estimation of θ_j suffered disproportionately. In each condition, estimation of θ_j seemed reasonable for all simulees except those with the most extreme values of θ_j .

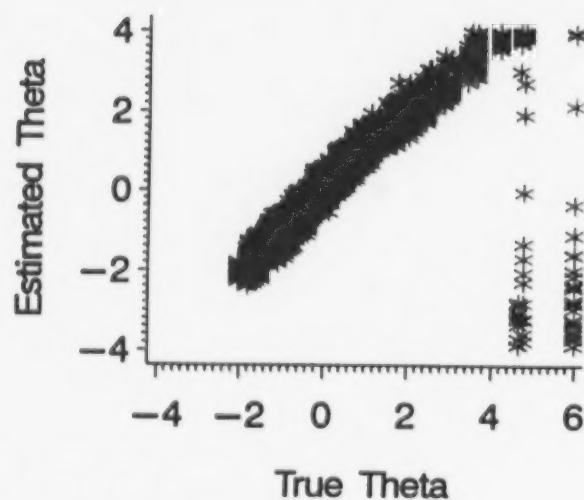
Difficulty in estimating θ_j for individuals who fail to agree with any items has been broached in both maximum likelihood (Roberts, 1995; Roberts & Laughlin, 1996a, 1996b) and EAP estimation contexts (Roberts, Donoghue & Laughlin, 1998, 1999). To assess the impact that extreme response patterns had in the conditions portrayed in Figure 13, the scatterplots were reconstructed after excluding any simulees whose average item response was less than .25. Given that there were 20 items on a 0-5 point response scale in which 5 represented strong agreement, those simulees below this cutoff did not strongly agree with any item on the test. The reconstructed scatterplots are displayed in Figure 14. Clearly, the estimation of θ_j improved substantially when those simulees with extreme response patterns were eliminated. When the average RMSE was calculated for the data from each panel in Figure 14, it ranged from .189 to .204. This level of error is similar to that obtained in the other simulated conditions.

The post-hoc findings portrayed in Figures 13 and 14 are conceptually straightforward. Those

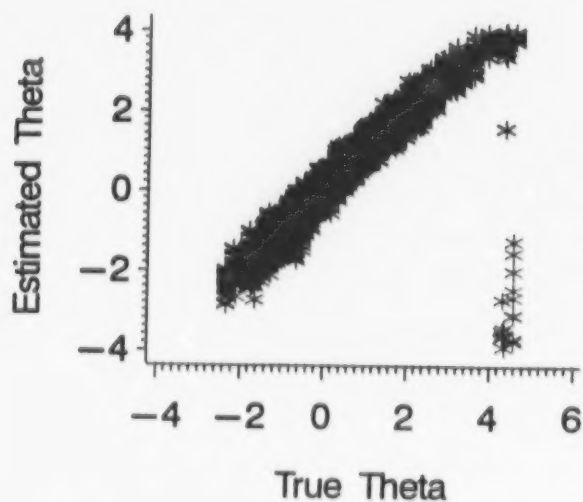
Negatively Skewed Generator
with Normal Prior and $N=500$



Positively Skewed Generator
with Normal Prior and $N=750$



Positively Skewed Generator
with Normal Prior and $N=1000$



Positively Skewed Generator
with Normal Prior and $N=2000$

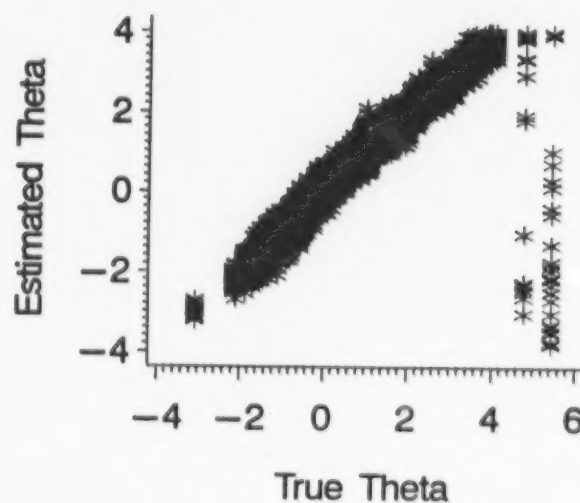
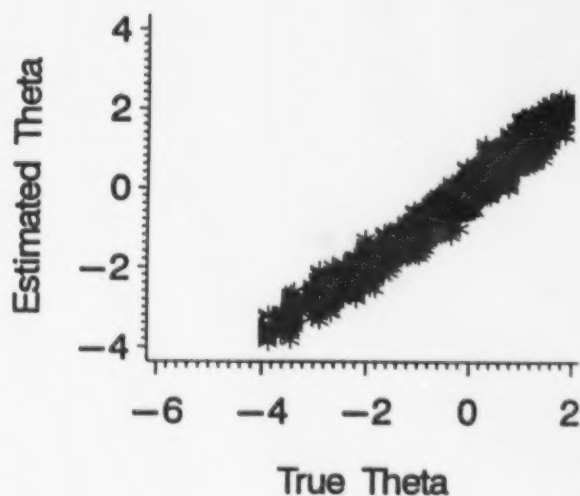
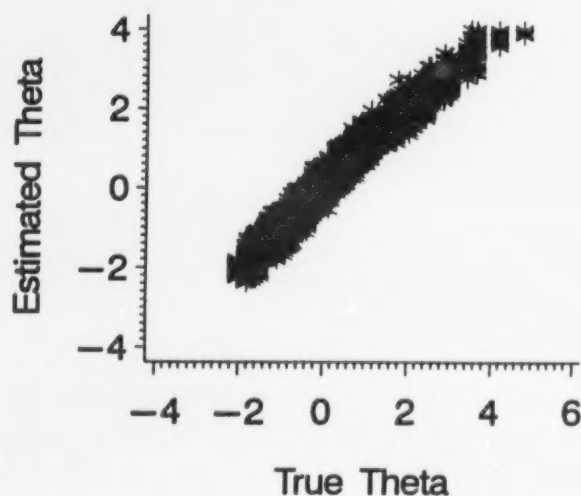


Figure 13. EAP estimates of θ_j versus true θ_j values for conditions which produced the most estimation error. The largest estimation errors were encountered in skewed tails of the true θ_j distributions.

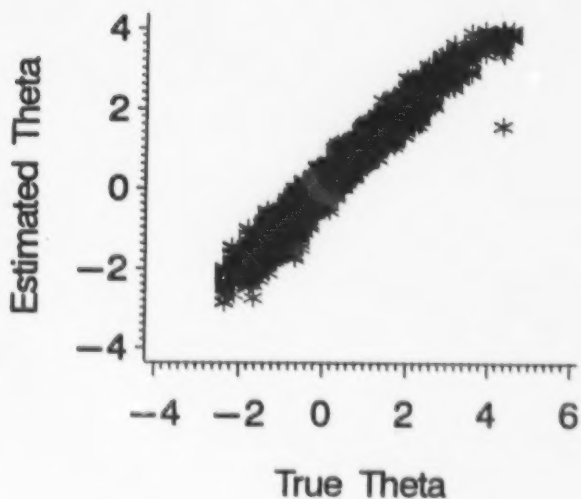
Negatively Skewed Generator
with Normal Prior and $N=500$



Positively Skewed Generator
with Normal Prior and $N=750$



Positively Skewed Generator
with Normal Prior and $N=1000$



Positively Skewed Generator
with Normal Prior and $N=2000$

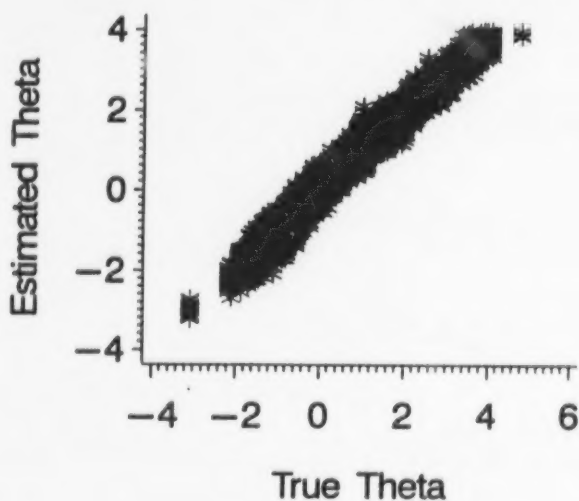


Figure 14. EAP estimates of θ_i versus true θ_i values for conditions which produced the most estimation error. Points corresponding to the most extreme response patterns have been censored.

individuals with nonextreme true θ_j values will generally have nonextreme response patterns. The information contained in these nonextreme response patterns will dominate the influence of the prior distribution when estimating θ_j values. Therefore, if reasonable estimates of item parameters are available, then relatively accurate θ_j estimates can be obtained. In contrast, when individuals are located far from any of the test items, they will, in all likelihood, fail to endorse any item to an appreciable extent. Moreover, the amount of information provided by the item responses will be minimal for these individuals. Under these circumstances, the observed data will provide little guidance as to the locations of these persons on the latent continuum. In such cases, the prior distribution will provide the bulk of information about the most appropriate location for these individuals. When that prior distribution differs markedly from the true distribution of θ_j , then those individuals with extreme response patterns will have relatively inaccurate estimates of θ_j due to the dominant influence of the discrepant prior distribution.

Discussion

The recovery results reported above suggest the MML item parameter estimates are relatively robust to discrepancies between the true distribution and prior distribution for θ_j . This implies that one can generally use a normal prior distribution when one is unsure of the true distribution of θ_j , and that the impact of this strategy on the accuracy of the resulting item parameter estimates will be modest at most. Moreover, even when the accuracy of item parameter estimates is slightly degraded by the discrepancy between prior and true θ_j distributions, the impact of this discrepancy on the accuracy of the estimated item response function will generally be even smaller. Although small improvements in the response function estimate can be gained by utilizing a matching prior distribution when one can approximate the true distribution of θ_j , this improvement is negligible relative to that which can be obtained by increasing the number of quadrature points used in the integration process to 20 or more. The accuracy benefits of correctly matching the prior and true distributions is also smaller than the increase in accuracy provided by larger sample sizes. These sensitivity results for the GGUM are similar to those found in past studies of other cumulative item response theory models (Bock & Aitkin, 1981; Seong, 1990; Bartholomew, 1988), and this convergence of findings across very different models provides more confidence about the general applicability of the MML procedure.

These results also reinforce the notion that MML estimation of item parameters in the GGUM improves with sample size, although large samples of up to 1000 subjects may be required before any diminishing returns on accuracy are noticeable. Nonetheless, the accuracy achieved before diminishing returns are encountered may still be acceptable in an absolute sense. This is especially true if one is concerned more with the accurate estimation of the item response function rather than the accuracy of particular item parameter estimates.

Another interesting finding is that EAP estimates of person parameters appear to be more sensitive to the prior distribution assumption than are MML estimates of item parameters. However, except for estimates of those θ_j values associated with extreme response patterns, the level of error seems to be generally acceptable. It may seem odd that the prior distribution of θ_j

should have so little an impact on the majority of θ_j estimates. However, the EAP method is a Bayesian technique. Therefore, given adequate estimates of item parameters, the impact of the prior distribution should decrease as the number of (informative) item responses increases. Therefore, the acceptable accuracy of EAP estimates when the prior and true distributions of θ_j are discrepant is primarily due to the previously mentioned robustness of item parameter estimates derived from the MML technique.

One exception to the general acceptability of EAP estimates in the GGUM occurs in the case where an individual fails to agree with any test items. In this case, the responses contain little information about the individual's location on the continuum, and thus, the impact of the prior distribution on the resulting estimate is very strong. If that prior correctly matches the true distribution of θ_j , then the resulting estimate may be quite accurate. However, large inaccuracies can result when the prior distribution does not adequately match the true distribution. For this reason, it seems reasonable to score only those individuals who exhibit some minimal level of agreement with at least one item. This strategy has been recommended in other applications of the model (Roberts, 1995; Roberts & Laughlin, 1996a 1996b; Roberts, Donoghue & Laughlin, 1998, 1999). We should also note that, theoretically speaking, this situation would only arise when an individual's location on the latent continuum was far from most, if not all, of the item locations for a given test. Therefore, this problem is at least partially under the control of the test developer, and items can be constructed to minimize the probability that this situation will occur.

Finally, the results of this study suggest that 20 quadrature points provides an acceptable level of numerical precision when calculating either MML estimates of item parameters or EAP estimates of person parameters. However, one must remember that the item responses simulated in this study conformed perfectly to the GGUM. In applications with real data, a more cautious perspective may be warranted, and thus, one may want to use more quadrature points.

Educational Importance

The impact of item response theory models for the measurement of ability in large scale educational testing situations has been enormous during the last 20 years. Item response models have led to scientifically sound approaches to assess differential item functioning (DIF) and to equate tests containing different items. They have also allowed for the development of sample free item banks and efficient computer adaptive testing (CAT) procedures. The family of models represented by the GGUM holds the same potential for other latent constructs that are important to educational researchers such as attitudes toward academic subjects, attitudes toward higher education, preferences for alternative teaching styles, etc. The GGUM could provide useful information in large scale educational testing situations and offer important insights about DIF, questionnaire equating, item banking and CAT in the attitude/preference domain similar to that provided by cumulative item response theory models in the ability domain. However, these insights can only be achieved to the extent that a reliable, accurate means of estimating GGUM parameters exists. The current study provides evidence that MML estimates of GGUM item parameters are robust to differences between the true distribution and the prior distribution for θ_j .

- a situation which is likely to occur to at least a moderate degree in practice. Moreover, this study also suggests that EAP estimates of person parameters in the GGUM are also reasonable in the face of the same discrepancies whenever response patterns are not extreme. The robustness of these estimation methods should help promote the use of the GGUM as a viable model for applied researchers, which, in turn, will help secure the benefits of item response theory models in the attitude/preference arena.

References

- Andrich, D. (1996). A general hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, 49, 347-365.
- Bartholomew, D. J. (1988). The sensitivity of latent trait analysis to choice of prior distribution. *British Journal of Mathematical and Statistical Psychology*, 41, 101-107.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement*, 6, 249-260.
- Linn, R. L., Levine, M. V., Hastings, C. N., & Wardrop, J. L. (1981). Item bias in a test of reading comprehension. *Applied Psychological Measurement*, 5, 159-173.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-176.
- Piessens, R., deDoncker-Kapenga, E., Uberhuber, C., Kahaner, D. (1983). *QUADPACK: A subroutine package for automatic integration*. NY: Springer-Verlag.
- Ramberg, J. D., Tadikamalla, P. R., Dudewicz, E. J., & Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21(2), 201-214.
- Roberts, J. S. (1995). Item response theory approaches to attitude measurement. (Doctoral dissertation, University of South Carolina, Columbia, 1995). *Dissertation Abstracts International*, 56, 7089B.

- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (1998). *The generalized graded unfolding model: A general parametric item response model for unfolding graded responses (RR-98-32)*. Princeton, N.J.: Educational Testing Service.
- Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (1999). A general model for unfolding unidimensional polytomous responses using item response theory. *Applied Psychological Measurement*, in press.
- Roberts, J. S., & Laughlin, J. E. (1996a). *The graded unfolding model: A unidimensional item response model for unfolding graded responses (RR-96-16)*. Princeton, NJ: Educational Testing Service.
- Roberts, J. S. & Laughlin, J. E. (1996b). A unidimensional item response model for unfolding responses from a graded disagree-agree response scale. *Applied Psychological Measurement*, 20, 231-255.
- Roberts, J. S., Laughlin, J. E., & Wedell, D. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, 59, 211-233.
- Roberts, J. S., Wedell, D. H., & Laughlin, J. E. (April, 1998). *Heightened sensitivity of Likert attitude scales to restriction of sample range*. Paper presented at the 61st Annual Meeting of the American Educational Research Association, San Diego, California.
- Seong, T. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*, 14, 299-311.
- van Schuur, W. H., & Kiers, H. A. L. (1994). Why factor analysis is often the incorrect model for analyzing bipolar concepts, and what model can be used instead. *Applied Psychological Measurement*, 18, 97-110.